
MOVING FROM TRADITIONAL DATA WAREHOUSE TO ENTERPRISE DATA MANAGEMENT: A CASE STUDY

Amit Pandey, Robert Morris University, axpst29@mail.rmu.edu
Sushma Mishra, Robert Morris University, mishra@rmu.edu

ABSTRACT

In the era of big data, organizations today rely of huge quantity of data from diverse sources and need to integrate this data in a speedy manner to gain any strategic advantage out of the data. Data warehouse is becoming increasingly popular in organizations due to the need for enterprises to gather all of their data in a single place for in-depth analysis and also to segregate such analytical work form on-line transaction processing systems. In carrying out this qualitative case study, researchers examined the limitation and issues of current data warehousing system architecture of the financial institution. The researchers discuss the benefits of migration from tradition data warehouse to enterprise data management and present the new architecture. Contributions are noted and conclusions drawn.

Keywords: Data warehouse, Enterprise data management, OLAP, OLTP, Case study

INTRODUCTION

Business today rely on various data sources not only for integrating its business processes in an operational sense but also for gaining advantage by analytically using information against competition. Combining various departmental databases such as sales, finance and human resource and coming up with integrated single repository of data for centralized decision making is not uncommon these days. In such circumstances, combining heterogeneous data sources in way such that currency and quality of data is not compromised is a challenge in itself. Research as well as business is plagued with problems arising from such integration efforts. A common phenomenon being witnessed by organizations dealing with such business requirements is implementation of data warehouses at enterprise level.

Data warehousing systems provide a platform such that information from operational data sources can be integrated into a central repository. Such a centralized source of data provides the ability for analysis and mining of the integrated information [3]. Data warehouse is becoming increasingly popular in organizations due to the need for enterprises to gather all of their data in a single place for in-depth analysis and also to segregate such analytical work form on-line transaction processing (OLTP) systems [15]. Currently data warehouses are used primarily in following areas [5]: On-Line analytical processing (OLAP) to support decision making, decision support systems, data mining for knowledge discovery process and database marketing to provide customized services to individual buyers.

In the era of big data, organizations today rely of huge quantity of data from diverse sources and need to integrate this data in a speedy manner to gain any strategic advantage out of the data. The argument of this study is that complexity of data management and usage in business environment today calls for enterprise data management planning. Disparate data warehouses today are unable to provide the flexibility and agility required for quick decision-making. Following the introduction, a brief literature review in this area is provided. The organizational context, where this case study was conducted, is introduced along with current issues with data warehouse in the firm. The section following this discusses the step towards enterprise management of data and benefits associated with it. Discussions are presented, future research directions are laid out and conclusions are drawn.

LITERATURE REVIEW

The need for creating a central repository of data for decision making purposes makes organizations consider the possibility of coming up with a data warehouse. Small databases lying in various functional areas of organizations are integrated thorough such development process. There can be alternate approach to integration of underlying databases. Pan and Vina [9] suggest such architecture based on two emerging

paradigms: web services and enterprise information integration (EII). The proposed architecture is claimed to be better than data warehouse approach on various parameters [9]: real time integration, less expensive, maintaining autonomous sources of data and creation of web account aggregation and business watch applications. On the other hand, real time access of information in this architecture reduces the performance of the overall system and maintaining such a system from a single framework is tedious.

Data in data warehouse often captures various aspects of business processes such as manufacturing, distribution, sales and marketing reflecting success and failure of various strategies and measures [6]. There has to be an appropriate data modeling approach for creation of data warehouses [6]. For the purpose of analysis, it is not only important to have all the data in one place but also crucial to know the sources of data. Tracing data warehouse source is called the lineage issue of data warehouse [3]. Such lineage tracing in a data warehousing environment has several benefits and applications, including in-depth data analysis and data mining, authorization management, view update and efficient warehouse recovery.

Data is generally cleaned before importing into a data warehouse, using process of transformations such as extraction-transform-load (ETL). Such transformations vary from simple algebraic operations to complex procedural code. The research challenge lies in realizing the ideal data warehouse architecture or create a warehouse specification language, rule capabilities, wrapper/monitor interfaces and appropriate algorithms. Such mechanisms would allow organizations to integrate various data sources and relevant change detection mechanism automatically [15].

DATA INTERGRATION IN DATA WAREHOUSE

Data integration can be defined as the process of “combining data residing at different sources and providing the user with a unified view of these data” [7]. Data warehousing has become popular in business environment where high performance query processing and data analysis is critical. Also, this is important when information sources are expensive, disjoint and access becomes occasionally access becomes unavailable and network delays cause cost [10]. Data warehouse can be effectively employed in an environment that is distributed but closed. In such business units, there are moderate numbers of distributed sources that act independently by trusted owners with joint cooperative goal [10]. The potential benefits of data warehousing are when the warehouse is used in the redesign of business processes and to support strategic business objectives [14]. Improved decision making usually results from the better information available from a data warehouse.

Data warehouses are generally designed around one of the two plans. Some data warehouses are designed with a set of data marts that partition the data warehouse into smaller, focused databases tailored to the information needs of a subset of users. Other data warehouses provide a single repository that gives the users a very wide range of data. Business units that accessed their data through data marts were usually more successful than units that accessed it through a single repository [2]. Some of the other factors that are important for success of data warehouse projects are [2]:

- The attitude of a unit’s leader affects all the factors leading to warehouse acceptance.
- Selecting the tools that will be available to the users is important. For user acceptance of the data warehouse. Restrictive tools which need high level of expertise to use are not welcome by the users.
- The relevance of the task to the organization, the degree to which a technology supports a task, and the degree to which users understand the task fit can influence the acceptance of the technology.
- Enough technical knowledge and expertise by users. A lack of knowledge of a technology can lead to difficulties using the technology, or even abandonment of the technology.
- There should experts within the business units who can acquire knowledge within the unit about data warehouse.

ORGANIZATIONAL CONTEXT: FIRM ABC AND DATA WAREHOUSE

The firm “ABC” is a financial institution located at western Pennsylvania. The firm deployed its first data warehouse system ten years back. Since the deployment of current data warehouse system, the firm has introduced multiple business strategies to meet the demand of its customer needs. The firm has gone through multiple re-organization processes since then. The IT steering committee reviewing the data warehouse architecture in the organization recommend that the current data warehouse system is not equipped to deal with fast changing data and business needs of the firm. The committee report suggests that the current data warehouse poses various challenges such as migration of data from legacy systems, maintenance of data quality, system management, and user education and training.

Figure 1 shows the current data warehouse architecture at the firm. The 3-layer data warehouse architecture is deployed in this firm. The first layer is operational layer that is consist of OLTP database, The second layer is data repository layer that consist of common data repository and OLTP databases and last layer is report layers.

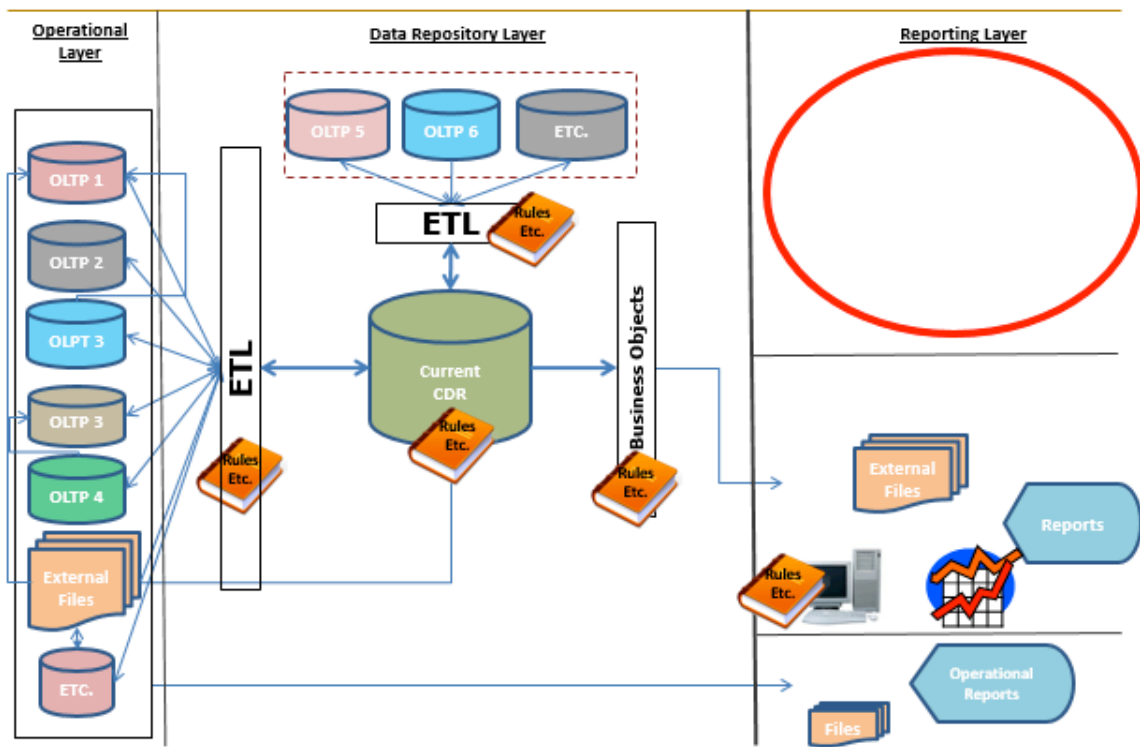


Figure 1: The current data warehouse architecture at firm ABC

The integration between operational layer and data repository layer occurs through ETL process. In Extract, Transform and Load (ETL) process, data are extracted from the OLTP databases and other sources of data such as xml file, plain text files and excel files and transformed and loaded using customer-written rules and software into CDR databases.

The integration between data repository layer and reporting layer occurs via Business Objects (BO). BO is reporting and querying tool used to provide access to the data warehouse. BO is a business intelligence software supported by vendor SAP. Business intelligence (BI) systems main objective is to enable interactive access to data, to enable manipulation of data, and to give business managers and analysts the ability to conduct appropriate analysis in a timely manner [11].

The IT steering committee at the firm identified the following important issues with its current data warehouse system.

Information in application silos

Each business unit in this firm has its own data mart. The customer information related data are stored in excel sheet, Microsoft access database and other databases. The data stored in multiple places cause duplication, discrepancy and creates application silos.

Accent [1] argues that “Application silos can either be created by the separation of data among systems, different application sets with their own data set, or by a combination thereof. Two different business units of an enterprise can have the same application installed however, their data separation can create silos that lead to loss of business or lack of efficiency (p.2)”. Though these silos might help in quick querying capabilities, the efficiency at overall enterprise is compromised due to redundancy. The data management plan clearly is not addressing end-to-end need of the firm in a holistic manner.

Operationally Focused

Operational systems are generally application specific. The operationally focused data are generally partial and non-integrated with other systems hence there is an issue with data redundancy. As described in Figure 1, for the operational reporting, there is a direct access between operational layer and reporting layer to access the data for day-to-day reporting purpose. The operational focus of data helps in efficiency of the processes however it could lead to myopic representation of information from enterprise level perspective.

Operational dependencies on common data repository

The common data repository (CDR) is tightly coupled with OLTP systems. Though CDR system is data warehouse system, the operational data store dependencies brings enormous challenge for online analytical processing systems.

Lack of metadata management

The issues of Meta data management for operational data stores do not significantly differ from the issues of Meta data management in the data warehousing. Meta data relevant for data warehousing are also relevant for application integration based on operational data stores and vice versa [17]. Data warehousing actually needs significant Meta data awareness [18].

Multiple “truths” of data

Data are sourced from multiple independent operational “legacy” system and external data provides to CDR system. Since each business unit used their own data marts, the data duplication issues and identifying up-to-date data are challenge with current data warehouse systems.

Departmental Analysis

The data analytics is focused on each business unit level or department level. The departmental analysis helps the firm to monitor day-to-day, weekly and monthly reports and activity for each department. There is however a significant disconnect between unit level reports and an integrated overall enterprise level report. There is lack of enterprise view of reports that could align with the firm business strategies and help senior executives in decision-making process.

The data services team within information technology department worked with other business units and departments to hash out the current data need for the reporting and decision making purposes. The table 1 present the current (as-is) state of the data warehouse system and desire (to-be) state of the data warehouse systems.

Table 1: Strategic goals for data management (AS-IS STATE / TO-BE STATE)

AS-IS STATE	TO-BE STATE
Information in application silos	Information aligned/ represented in macro processes
Operationally focused	Analytics, Drill-thru Performance management
Operational Dependencies on CDR	Operational/Analytical independence
Lack of Metadata Management	Metadata Management Methodology
Multiple “Truths” of data	Single “Truth” of data
Departmental analysis	Enterprise process focused

Despite the potential benefits of data warehousing and OLAP tools, data warehouse implementation projects are difficult to use and fail to realize the benefits [4]. It needs proper planning on part of management for success of such projects. Research in this area identifies some factors for success of such projects. Gorla [4] claims that OLAP tools should be designed considering alignment of IT strategy with business strategy, for success of data warehouse projects. Vendors should design tools using better GUI and drag and drop technologies. Better physical data warehouse design is helpful in improving the performance of the analytical tools for decision-making. Improved currency of the data warehouse with frequent updates also helps in establishing the importance of a data warehouse in an organization. All the factors mentioned above combine and increase the perceived benefit of data warehouse and facilitate better adoption. Watson and Wixom [18] found that system quality and data quality play an important role in success of a data warehouse project. High systems and data quality ensures better success of the project and vice versa. Shin [12] conducted a study that confirmed that the system was critical in serving a variety of tasks from routine work to complex planning and decision-making. Users had a high regard for the role of the data warehouse in enhancing their work productivity.

DEVELOPING EDM AT ABC FOR DATA INTEGRATION

The IT steering committee at ABC in its strategic planning endeavors has decided to change its fundamental approach to data usage. As a strategic step, firm ABC has decided to migrate from traditional data warehouse techniques to enterprise wide data management solutions. The organization is in the process of this conversion.

Enterprise data management (EDM) provides a single view of the truth, unique reference data and a unified data quality framework to integrate, validate and migrate data. EDM enables continuous monitoring of the quality of data and establishes standards across the enterprise data lifecycle [13]. Mongoven [8] argues that a comprehensive EDM solution includes “capabilities in profiling, cleansing and monitoring to improve data quality; metadata (data about data) to ensure consistent definitions and usage; data governance to provide ongoing oversight and enforcement of standards and procedures; retention and security to support accessible and secure data; and master data management. Improvements in these areas can help produce the kind of data your business can rely on – from the beginning” (p. 1).

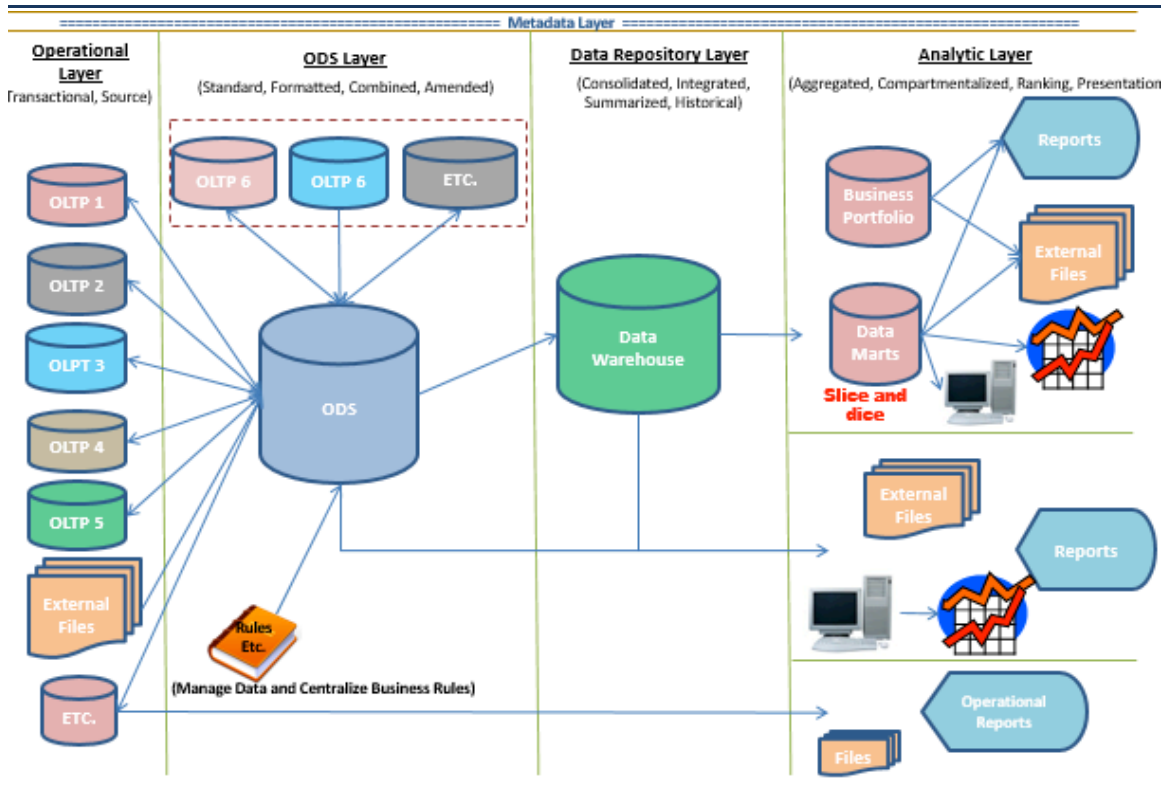


Figure 2 : Proposed data warehouse architecture at firm ABC

Utilizing EDM approach , the data services team proposed new data warehouse architecture at the firm ABC. The new data warehouse architecture consist of five layers. Operational Layer, ODS Layer, Data Repository Layer , Analytics Layer and Metadata layer. Operational Layer will consist of transactional data and would be source data for ODS layer. In ODS layer, the data would be formatted, amended, standardized and managed by centralized business rules. The ODS layer will be a source data point for data repository layer. In data repository layer historical business data and weekly, monthly and daily business data will be consolidated, integrated and summarized. The analytic layer would be responsible for reporting and analyzing the data for decision making. In this layer, the business data would be stored in data marts as per the need. The slice and dice mechanism would be utilized to understand and examine the data from different viewpoints. The segregation of analytical level from operational level is useful and more efficient in understanding the nature of data for operational as well as strategic decision purposes.

DISCUSSIONS

Data warehousing and functional data sources share certain similarities and there are significant synergies in development and such opportunities should be exploited [17]. Research in these area points that there are methodological reuses potential for project justification and systems development and any expertise in this area should be exploited by organizations for optimization of resources [17]. The greatest potential benefits of data warehousing are when the warehouse is used in the redesign of business processes and to support strategic business objectives [14]. Improved decision making usually results from the better information available from a data warehouse.

The proposed enterprise data management approach would provide consolidated view of data eliminating the need to gather data from multiple input sources; reduce time for reconciliements and report development for end users; reduce risk from manual exposure calculations due to service availability and expand user access to data. With “more is better than less” data storage policies, lower cost of data storage, low cost processing power per data unit, we are gradually moving towards more analytical processing at a faster pace. On this context, enterprise data management is the requirement of this financial institution and a step towards implementing it is the right step towards strategic focus of this firm.

This study contributes in multiple ways: First, there are few studies in this domain acknowledging the benefits of EDM over traditional data warehouses. Second, this study one of the few attempts that document this comparison and argues for EDM over tradition data warehouse and last but not least, it provides a good reference guide for the practitioners struggling with traditional data warehouse issues. This study is not without limitations. This situation described here is unique to this firm where the study was conducted. As in the case of any single case study, generalizability of the findings is questionable. In case of any qualitative case study, certain probability of researcher bias creeping in is always present. This study can lead to several future studies in this area. A quantitative study about benefits of EDM of disparate data warehouse is called for. Also at an enterprise level, a study that compares the efficiency and effectiveness ratios of the firm, in case of disparate and enterprise data management, would also provide interesting insight into this phenomenon.

CONCLUSIONS

Organizations have used data warehouses over the years to integrate data from multiple sources and use the unified repository for data mining to identify relationships/patterns in the data. With complexity and volume of data increasing over time, a single stand-alone data warehouse is not able to provide the strategic benefits required by the firm ABC. This firm is moving towards an enterprise wide data management solution that it can sustain in the future to meet its analytical as well as operational needs for decision making. The case study provided here emphasizes that there is value in addressing the data needs of the firm in a holistic manner.

REFERENCES

1. Accent, V. (2006). Do application silos make sense in today's enterprise? from <http://computer.financialexpress.com/20060501/technology02.shtml> accessed on 3/24/2014.
2. Chenoweth, T., Corral, K. and Demirkan, H. (2006). Seven Key Interventions for Data warehouse Success. *Communications of the ACM*, 49(1), pp. 114-119
3. Cui, Y. and Widom, J. (2003). Lineage tracing for general data warehouse transformations. *VLDB Journal*, 12:41-58/ Digital Object Identifier, pp. 41-58
4. Gorla, N. (2003). Features to Consider in a Data Warehousing System, *Communications of the ACM*, November, Vol. 46, No. 11, pp. 111-115
5. Gray, P. and Watson, H. (1998). Present and Future Directions in Data Warehousing. *The Database for Advances in Information Systems*, Vol. 29, No. 3, pp. 83-90
6. Jukic, N. (2006). Modeling Strategies and Alternatives for Data Warehousing Projects. *Communications of the ACM*, April, Vol. 46, No. 4, pp. 83-88
7. Lenzerini, M. (2002). *Data Integration: A Theoretical Perspective*. ACM PODS 2002, June 3-6, Madison, Wisconsin, USA
8. Mongoven, T. (2013). Enterprise Data Management: Rapid Evolution. from http://www.deloitte.com/view/en_US/us/Services/consulting/technology-consulting/7e4f50b1e37210VgnVCM100000ba42f00aRCRD.htm accessed on 4/1/2014.
9. Pan, A. and Vina, A. (2004). An Alternative Architecture for Financial Data Integration. *Communications of the ACM*, Vol. 47, No. 5, pp. 37-40
10. Rundensteiner, E., Koeller, A. and Zhang, X. (2000). Maintaining Data Warehouses over Changing Information Sources. *Communications of the ACM*, June, Vol. 43, No. 6, pp. 57-62
11. Sap.com (2014). BusinessObjects. From <http://www.sap.com/pc/analytics/business-intelligence.html> accessed on 5/10/2014.
12. Shin, B. (2003). An Exploratory Investigation of System Success Factors in Data Warehousing. *Journal of the Association for Information Systems*, 4, pp. 141-170
13. Sivaprakasam, S. R. (2010). Enterprise data management A comprehensive data approach for CSPs. from www.infosys.com/industries/communication-services/white-papers/Documents/comprehensive-data-approach.pdf access on 4/12/2014.
14. Watson, H. and Haley, B. (1998). Managerial Considerations. *Communications of the ACM*, 41(9), pp. 32-37

15. Widom, J. (1995). Research Problems in Data Warehousing. CIKM '95, Baltimore MD USA, pp. 25-30
16. Winter, R. (1999). Data Warehousing beyond Tools and Data: Justification, Organization, and Structured Development of Data Warehousing Applications, in: Abramowicz, W., Orłowska M.E. (Eds.): Business Information Systems – Proc. BIS'99, Practical Sessions, pp. 197-208.
17. Winter, R. (2001). The Current and Future Role of Data Warehousing in Corporate Application Architecture. Proceedings of the 34th Hawaii International Conference on System Sciences, Hawaii, USA
18. Wixom, B. and Watson, H. (2001). An Empirical Investigation of the Factors Affecting Data warehousing Success. MIS Quarterly, Vol. 25, No. 1, pp. 17-41